

DIAL: A Platform for real-time Laboratory Surveillance

Shamir N Mukhi, PhD, PEng^{1,2}, Jennifer May-Hadford, MPH^{2,3},
Sabrina Plitt, PhD^{2,3}, Jutta Preiksaitis, MD⁴, Bonita Lee, MD, MSc³

¹Canadian Network for Public Health Intelligence,
²Public Health Agency of Canada,
³Provincial Laboratories for Public Health Alberta,
⁴University of Alberta

Abstract

Laboratory information systems fulfill many of the requirements for individual result management within a public health laboratory. However, access to the systems by data users, timely data extraction, integration, and data analysis are difficult tasks. These difficulties are further complicated by often having multiple laboratory results for specific analytes or related analytes per specimen tested as part of complex laboratory algorithms requiring specialized expertise for result interpretation. We describe DIAL, (Data Integration for Alberta Laboratories), a platform allowing laboratory data to be extracted, interpreted, collated and analyzed in near real-time using secure web based technology, which is adapted from CNPHI's Canadian Early Warning System (CEWS) technology. The development of DIAL represents a major technical advancement in the public health information management domain, building capacity for laboratory based surveillance.

Keywords: Laboratory, Surveillance, Informatics, Analysis, Epidemiology, Pandemic

Introduction

Public health laboratories are the front-line of public health. The ability of the public health system to respond to emerging public health challenges is partly dependent on how effectively public health laboratories, create, integrate and communicate testing results [1]. The recent H1N1 pandemic has confirmed the important need for timely access to comprehensive laboratory testing data in order to monitor trends and rapidly make informed public health policy decisions. Historically, it has been difficult to access laboratory data for surveillance purposes as Laboratory Information Systems (LIS) focus mainly on data handling for laboratory testing including the pre-analytical, analytical and post-analytical phases relating to clinical case management [2]. Moreover, the ongoing identification of new pathogens associated with clinical syndromes, technological advancements in diagnostic assays and the use of complex test algorithms to improve diagnostic sensitivity and specificity have made it difficult to interpret and present laboratory data in a format that is useful for surveillance and public health purposes.

In 2007, the Provincial Laboratory for Public Health (ProvLab) in Alberta and the Canadian Network for Public Health Intelligence (CNPHI) [3] partnered to create DIAL, Data Integration

for Alberta Laboratories, a platform allowing laboratory data to be extracted, interpreted, collated and analyzed in near real-time using secure web based technology, which is adapted from CNPHI's Canadian Early Warning System (CEWS) technology [4].

Alberta is a province in Canada with a population of about 3.7 Million. Alberta's geography is diverse, ranging from rural farming communities to areas of natural resource extraction to highly urbanized centers. All health services are publicly provided and are managed and directed under a single health authority since April 2009. ProvLab provides a wide array of diagnostic tests for communicable diseases, in concert with regional laboratories, as well as specialized testing, reference services and laboratory support for provincial programs such as prenatal screening and outbreak investigations. ProvLab also provide diagnostics services for the Northern Territories including the Northwest Territories, Nunavut and parts of the Yukon. The diagnostic manual at ProvLab includes testing and characterization of viral, bacterial, fungal and parasitic pathogens from various sample types using traditional methods, advanced molecular techniques, and serological tests.[5]

Historically in ProvLab, laboratory data was accessed through rigorous extraction routines from LIS to a data management tool that enabled a simpler interaction with the data. However, customized queries and specialized tables were accomplished only by programmers and thus data was not easily accessible to medical, technical and administrative laboratory staff or public health practitioners outside the laboratory. Moreover, the raw laboratory data still required interpretation and transformation by laboratory experts to become clinically meaningful data for public health stakeholders and researchers. A solution that allowed front line users to access, present and analyze collated, integrated and interpreted laboratory information in real-time was needed.

We describe the solution to this need, the DIAL platform, focusing in depth on its various key components as displayed in Figure 1. The platform has been developed to work with any disease target tested within the laboratory, however, given the complexity of laboratory testing procedures for various targets, the platform needs to be adapted as needed. At its present state, the DIAL platform has been configured to analyze testing data for the detection of respiratory viruses, typing of methicillin resistant *Staphylococcus aureus* (MRSA), syphilis (serological assays and direct detection), and environmental water testing.

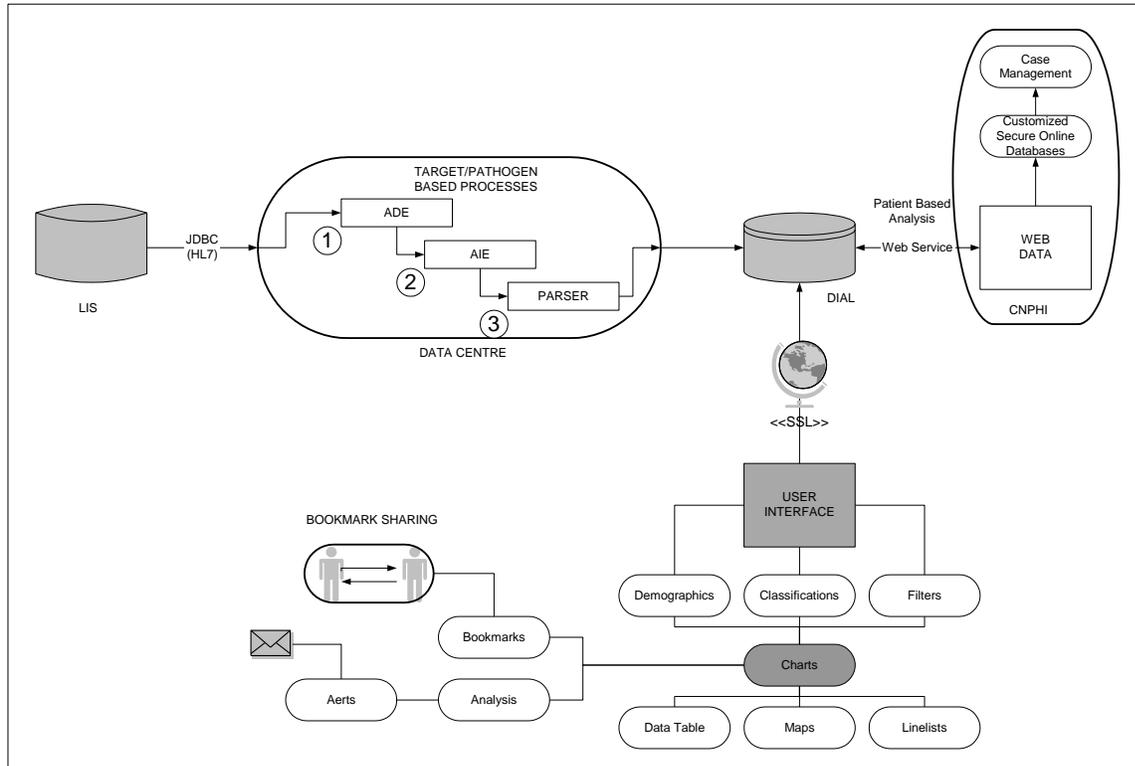


Figure 1: DIAL Platform Components

DIAL Platform Components

Technical Configuration

DIAL platform runs using JBoss server and Oracle database on a DELL R900 server using VMware virtualization technology. The DIAL platform requires three virtual servers including:

1. *Database:* Oracle database
2. *Data Centre:* data processing facility
3. *Application:* user interface

DIAL has proven to be a very stable system since its launch in 2009. DIAL's speed is not expected to decrease as the number of users are increased, nor is the stability expected to decline as more data or disease targets are added.

Data Acquisition

DIAL data is virtually real-time as laboratory information is automatically extracted using a proprietary extraction engine connected directly to the LIS twice daily. LIS data is organized by specimen and contains demographic and geographic information on patients and health providers as provided on requisitions accompanying the submitted specimens. The latter data is routinely entered manually as part of the pre-analytical testing process. Test requests for specific analytes

are entered into LIS using specific test codes and results of requested tests and supplementary or confirmatory tests, as defined by testing algorithms, are entered into the LIS electronically for tests performed on interfaced instruments or manually by technologists. The extraction and provision of LIS data to DIAL does not require any effort from laboratory personnel.

The specimen data from the LIS are automatically extracted into the data management tool running on a SQL server database allowing seamless connection over JDBC (Java Database Connectivity). The data is extracted using proprietary automated data extraction engine (ADE) which is a data extraction approach using one line per specimen from the LIS. The data is extracted using standard connection to the relational database which can be adapted to various LIS platforms. All the relevant specimens data contained in the LIS are extracted into the DIAL database.

Currently, the ADE runs twice daily, however, it can be scheduled to execute more often if required. Once extracted, the laboratory test results for the extracted specimens go through an interpretation engine and are stored in an Oracle based database. DIAL is developed as a record based system where a record can be a specimen or a patient depending on the source data. Given the complexities of compiling specimen-based data into patient-based information, especially for infectious diseases with the possibility of re-infection events occurring at variable intervals, DIAL currently only supports specimen level reporting and analysis.

DIAL does not allow modification of extracted data, thus, complying with the requirement that the LIS is the “owner” of the data and changes must be enacted in the LIS. Automatic electronic updates directly from a LIS will outperform, both in time and completeness, other laboratory surveillance systems that rely on manual data entry, human interpretation and communication [6,7].

Security

DIAL is accessed via a secure, password protected web based platform using SSL (secure socket layer) technology hosted at the ProvLab facilities. DIAL’s provision over the web allows remote access to the application. There is no need for hardware or software enhancements by users. Individuals applying for system access are evaluated by a ProvLab team to assess their needs. Data access for external users is provided in accordance with health information access policies defined by health partners and stakeholders. If DIAL access is granted, a user agreement is signed outlining the ownership and usage of the data. Current users include medical officers of health, communicable disease nurses, provincial and local epidemiologists, laboratory staff, and ProvLab laboratory personnel.

DIAL supports tiered access control using proprietary registration system. Access to data can be restricted by geography (e.g.: province, region), targets/pathogens (e.g.: respiratory, mrsa) and function (e.g.: trending, algorithms).

Interpretation engine

The detection of pathogens is often achieved by a variety of test methods which generate multiple test results for a sample. These test methods change overtime with technological advancement. For example, depending on the type of sample and the year of submission, various respiratory viruses could be tested using one or more of these methods: direct florescent antibody tests, traditional virus culture, rapid shell vial culture, in-house developed or commercial-based molecular diagnostic tests. These multiple test results generated for a sample need to be interpreted by laboratory experts to provide clinically meaningful data. Similarly, serological tests frequently include complicated algorithms with sensitive screening assays to be followed by confirmatory tests. The data generated by these multi-test algorithms is complex, requiring an enormous amount of time and efforts for data cleaning and interpretation, which presents a major challenge for the laboratory experts. To address this challenge, an automated interpretation engine (AIE) was developed with the content experts at ProvLab. The AIE comprises of target specific interpretation script that analyze raw laboratory results and classify each specimen based on set rules to provide interpreted laboratory data. Depending on the type of sample and tests performed, there may be multiple classifications per specimen, e.g., multiple respiratory viruses that are detected by various methods in a sample. With updates and changes in test algorithms over time, multiple interpretation algorithms may result in the same classification, e.g., influenza A can be detected by one or more methods reported using different test codes and result formats in LIS. Depending on the data needs, AIE can also be constructed to provide detailed laboratory technical information so that the testing data can be easily accessed. This could be used to evaluate test algorithms and assay performance characteristics to support laboratory quality initiatives. The AIE and its' associated classifications translate results from complex laboratory algorithms into clinically meaningful data forming the basis for data analysis in DIAL.

AIE, which is adapted from the smart engine technology [4], uses the following functions to process the data:

- Data receiver: Allows importing of multi-format data from flat files and/or databases. In its' current state, data is presented to the AIE in a tab delimited format.
- Data validation: Applies predefined logical rules to incoming datasets to filter any records that are missing key data elements. For example, a record must have a patient identifier.
- Data coder: Converts data values into standard analyzable codes. This is where the *classifications* are generated. A classification can be lowest level *interpretation* (code representing specific set of rules based on various test results) or a group of interpretations, or other classifications, thus providing a mechanism for a hierarchical structure.
- Data export: Facilitates exporting of data in various formats to facilitate parsing the data into appropriate database table(s).

Interrogation, Interactivity, Analysis, Reporting

DIAL data are accessible by users via an interactive web based platform. The user interface is intuitive, based on a simple point and click approach. The user interface built in Java technology (JSF - java server faces) is based on the following main components:

- **Classifications:** The fundamental unit of analysis in DIAL is the classification (or a set of classifications) assigned by the AIE to each specimen. Single or multiple classifications must be selected to create a “data series” for display. The portion of the interface that allows one to specify a classification within a target (e.g. influenza B positive amongst all respiratory samples in a defined time period) is organized in an expanding tree fashion with check boxes listing all possible classifications related to a given target. This enables users to seamlessly navigate through the tree and make appropriate selections.
- **Demographics:** The interface enables users to select specific settings for age, gender, geography (including provincial, regional, city and postal code) and other parameters, e.g., outbreak versus sporadic community-based specimens, enabling customized demographic, geographic and specimen type ranges for data analysis.
- **Chart Configuration:** DIAL interface enables plotting of various types of charts including line, bar, stack and pie charts. Special features such as a cumulative curve on line graphs can be added. Individual charts can be constructed based on each selected classification or in a combined format with a single chart displaying all classifications simultaneously. The interface provides a mechanism to customize the colors, axis, legends, series order, titles and markers.
- **Advanced Filters:** In addition to classification selection and user-defined time period, demographic and geographic settings, users are able to further filter records using pre-defined set of fields which could be built as simple pick lists or complex wildcard searches. The ability to filter records aids in the ability to stratify and select data, especially in data fields that allow a wide variety of content, for meaningful analysis and reporting. This reporting flexibility adds to the usability and acceptability of the system.
- **Rates:** Visualizing rates is an important aspect of any surveillance system. Specifically, the rate of the number of specimens within one classification groups in relation to the others through an interactive interface is of value for analyzing laboratory data. DIAL interface enables users to select up to two rate computations which are displayed as overlays on the main chart.
- **Bookmarks:** Chart specification and customization can take significant effort, thus, repeating the process on a routine basis can be time consuming and frustrating. To alleviate this, DIAL allows users to bookmark charts including all the settings, customizations and filters. These bookmarks can be shared and retrieved on a routine basis when generation of similar charts is desired.

- **Data Table:** The interface includes the ability to convert a chart into a summary data table that can be exported to a comma delimited format (easily viewed using Microsoft Excel). Further, each cell can be viewed as a detailed line list (that is, all fields of a given record). These line lists of records may be exported as a comma delimited file.
- **GIS Maps:** During the parsing stage (after extraction and interpretation), all specimens are assigned a geolocator using a specific algorithm that takes into account various variables including postal code, city and health region, as available. This geolocator is stored with the specimens in the DIAL database. These geolocators can be plotted on a GIS map (via Google Maps) as points or summarized at a regional level. These maps can be exported for inclusion in reports. The ability of DIAL to represent data on maps is valuable to communicate information and emphasize geographic relationships.
- **Analysis:** Matlab analytical package has been integrated into DIAL to enable advanced analysis of data series. Due to the potential complexity of using advanced analytical packages, the DIAL interface includes two options: a basic mode which includes pre-defined functions; and an advanced mode which enables custom scripting of functions.

Examples of DIAL applications

DIAL has become the core infrastructure supporting laboratory based surveillance and reporting for ProvLab. Below are some examples demonstrating DIAL's utility.

1. *Reporting during Pandemic H1N1:* DIAL was used to provide daily surveillance reports including rates and linelist data for circulating respiratory viruses, including influenza A by H-typing results, to the province of Alberta, the territories, Nunavut and Northwest Territories. Summary data was provided for national surveillance [8].
2. *Test Optimization:* Data provided by DIAL was used to assess the sensitivity and specificity of various test methods, including newly implemented molecular diagnostic tests specific for pandemic influenza H1N1 2009 during the first wave of the pandemic. Without DIAL this would have been unachievable due to intense direct hands-on time required for specimen processing and result generation by laboratory staff during the pandemic. By combining real-time data on circulating viruses from DIAL with knowledge of test characteristics, a model was able to be established which performs cost/quality analysis of different testing algorithms thus supporting optimized and cost-effective decision making with respect to the sequence of tests performed within the laboratory during the pandemic [9].
3. *MRSA Analysis:* ProvLab has characterized all first case MRSA isolates, (defined as first infection from patient within previous 12 months), forwarded by diagnostic laboratories in Alberta as part of an enhanced surveillance program since June 2005. MRSA strains are reported as one of the Canadian epidemic strain types (CMRSA1-10), based on PFGE and spa typing. We used DIAL to analyze the typing data for all MRSA isolates received by ProvLab between June 2005 and December 2008 [10]. DIAL can be used by external stakeholders to access and analyze MRSA typing data.

Limitations and Future directions

- Only ProvLab data is currently being utilized in the DIAL system. For many pathogens, ProvLab performs all testing within the province providing population-based data. For some markers, however, other laboratory partners within Alberta perform variable portions of the testing and negotiations and discussions are underway with respect to including their data on DIAL which will result in more comprehensive and integrated laboratory data within the province.
- The development of a method to convert specimen based data to patient based data is planned. This will allow a better determination of the incidence of infection as multiple samples from one individual will be rationalized by the system. The goal is to create a patient record within DIAL containing all the laboratory samples for every patient. Every Alberta resident has a unique public health number for provincial health insurance claims; this will be useful in the creation of a patient record.
- An aberration detection component of the system is planned. This functionality will be developed based on confidence based aberration interpretation framework [11], using Matlab and Java for implementation of the algorithms. A system for automated and configurable alerts will be setup so that notifications can be generated when predetermined thresholds in both time and space are exceeded enabling the early identification of outbreaks or clusters of infections.
- The AIE for different communicable disease targets, e.g., enteric pathogens, blood-borne pathogens, will be developed in the DIAL platform in the future.

Conclusion

Based on our literature survey and personal communications, other international jurisdictions [12-15] have also realized the importance of reporting summarized laboratory data for public health intelligence. It does not appear that any of the systems are identical, all having evolved to satisfy the needs of their local communities of practice. The processes used to import laboratory data are different, ranging from manual input to automatically downloaded laboratory files over the internet. The ability to perform custom queries of data appears to be similar in the systems reviewed; however the use of a global laboratory interpretation engine, a proprietary feature of the DIAL platform, appears to be unique. We believe that a complete and thorough systematic review of currently functional laboratory surveillance systems is needed in order to highlight best practices that could guide further development.

DIAL has been a valuable addition to communicable disease surveillance in Alberta. DIAL's ability to report real-time, interpreted laboratory data proved invaluable during the recent pandemic influenza outbreak when timely and accurate data was needed for reporting and decision making. The abilities of DIAL platform to extract data, clean the data, transform it using automated interpretation engines, integrated the data, and to display data using graphics, tables

and maps through interactive user interface further enhances the ability to interpret and communicate information to laboratory staff, laboratory stakeholders and policy makers.

Acknowledgements

The authors would like to acknowledge contributions of the CNPHI and ProvLab teams. Specifically, Dr Marie Louie, the acting medical director of the ProvLab, for her ongoing support, and Lin Yan, David Sanders and Tony Huynh for their assistance with the development of the platform.

Correspondence:

Shamir N Mukhi

shamir.nizar.mukhi@phac-aspc.gc.ca

References

- [1] Canadian Public Health Laboratory Network. Core Functions of Canadian Public Health Laboratories. Accessible at: http://www.cphln.ca/pdf/2004-09-14_CPHLN_Core_Functions_eng.pdf
- [2] Reference Chapter VI – Laboratory Information Systems in “Laboratory Medicine: A National Status Report”. Prepared by The Lewin Group for Division of Laboratory Systems, National Center for Preparedness, Detection, and Control of Infectious Diseases, Centers for Disease Control and Prevention. 2008.
- [3] Mukhi, S. N., Aramini J. and Kabani A. Contributing to communicable disease intelligence management in Canada, *Can J Infect Dis Med Microbiol.* 2007 November; 18(6): 353–356.
- [4] Aramini J. and Mukhi S. N. Canadian Application of Modern Surveillance Informatics. In: Lombardo J.S. and D. L. Buckridge Disease Surveillance: a public health informatics approach, pp. 315-328. John Wiley & Sons Inc. Publication.
- [5] Guide-to-service Version 4.3 – 2010 December 08. Provincial Laboratory for Public Health (Microbiology) and Medical Microbiology Laboratory, University of Alberta Hospital, Edmonton, Alberta. Available at: <http://www.provlab.ab.ca/guide-to-services.pdf>
- [6] Ward M, Brandsema P, Van Straten E, Bosman A. Electronic reporting improves timeliness and completeness of infectious disease notification, *The Netherlands*, 2003. *Euro Surveill.* 2005;10(1):pii=513 Available at: <http://www.eurosurveillance.org/viewarticle.aspx?articleid=513>
- [7] Nguyen T., Thorpe L. Makki H., Nostashair F., Benefits and Barriers to Electronic Laboratory Results Reporting for Notifiable Diseases: The New York City Department of Health and Mental Hygiene Experience. *Am J Public Health* v97(supplement_1): s142-s145
- [8] Mukhi, S.N., B. E. Lee, J. K. Preiksaitis, 2009: DIAL: An interactive web based platform for real-time interpretation and analysis of laboratory test results to support analysis and reporting during the H1N1 Swine Origin Influenza Virus (S-OIV) outbreak, AMMI-CACMID, 2009.
- [9] Lee BE, Louie M, Chamberlin B, Fenton J, Flegel M, Fonseca K, May-Hadford J, Plitt S, Tellier R, Mukhi S, Drews S. Optimization of respiratory virus test algorithms by cost and

- quality during the influenza (H1N1) 2009 pandemic using a secure web-based platform. 2010 AMMI Canada - CACMIID Annual conference, Edmonton, Canada, May 6-8, 2010.
- [10] Ferrato C, Chui L, Mukhi S, Drews S, Lovgren M, Yan L, Checkley S, Lee BE, Louie M. A web based platform for MRSA typing surveillance in Alberta. 2010 AMMI Canada - CACMIID Annual conference, Edmonton, Canada, May 6-8, 201
- [11] Mukhi S. A Confidence based aberration interpretation framework for outbreak conciliation. Online Journal of Public Health Informatics, ISSN 1947-2579, <http://ojphi.org> Vol.2, No. 1, 2010.
- [12] Rolfhamre P, Janson A, Arneborn M, Ekdahl K. SimiNet-2: Description of an internet-based surveillance system for communicable diseases in Sweden. Euro Surveill. 2006;11(5):pii=626 Available at: <http://www.eurosurveillance.org/viewarticle.aspx?articleid=626>
- [13] Faensen D., Claus H., Benzler J., Ammon A., Pfoch T., Breuer T., Krause G., SurvNet@RKI- A multistate electronic reporting system for communicable diseases; Eurosurveillance Vol.11 Issue 4-6 pp100-103.
- [14] Domeika M., Kligys G., Ivanauskiene, O., Mereckiene J., Bakasenas V., Morkunas B., Berescianskis D., Wahl T., Stenqvist K., Implementation of a national electronic reporting system in Lithuania; Eurosurveillance Vol.13 Issue 13 pp 1-6.
- [15] CIDR -Computerized infectious disease reporting, e-health 2005 presentation Accessed on April 27 at <http://www.hpsc.ie/hpsc/CIDR/Presentations/File,1113,en.pdf>.